Alignment of Deep Neural Representations and Radiomics for Improved Kidney Tumor Classification

I-Chun Tsai[1] GINAGIGO1220@GAPP.NTHU.EDU.TW
Hsi-Yu Ho[1] ARMANI113062581@GAPP.NTHU.EDU.TW
Huan-Yu Chen[1] RAY.CHEN1996@GMAIL.COM
Chun Chieh Weng[1] CHCHWENG@GMAIL.COM
Jing-Tong Tzeng[1] ROGER37890426@GMAIL.COM
Ching-Heng Lin[2] EPID@VGHTC.GOV.TW
Chi-Chun Lee[1] CCLEE@EE.NTHU.EDU.TW
Po-Chih Kuo[1] KUOPC@CS.NTHU.EDU.TW
[1] *National Tsing Hua University*

[2] *Taichung Veterans General Hospital*

Editors: Under Review for MIDL 2026

Abstract

Preoperative diagnosis of kidney tumors, including the crucial differentiation between benign and malignant lesions and the grading of malignant tumors (low-grade vs. high-grade), is vital for treatment planning but challenging with traditional imaging alone. This study proposes a novel Radiomics-Guided Convolutional Neural Network (RGCNN) model that aims to enhance kidney tumor classification accuracy by integrating Computed Tomography (CT) images with quantitative radiomic features. Data preprocessing involved isotropic resampling and Connected Component (CC3D) cropping on datasets from three different sources (KiTS19, TCGA-KIRC, VGHTC) to standardize the tumor region. The RGCNN model not only utilizes Traditional Radiomics (TR) features but also extracts CNN-Based Radiomics (CR) features from its convolutional layer feature maps. Experimental results show that combining both CR and TR outperforms standalone CNN or TR-only models. The combined model achieved an AUC of 0.82 and an F1-Score of 0.89 in the benign-versus-malignant classification task, and an AUC of 0.74 and an F1-Score of 0.64 in the more challenging low-grade versus high-grade classification task. This study validates the significant potential of RGCNN as a non-invasive, preoperative diagnostic aid.

Keywords: Kidney Tumor Classification, Radiomics, Convolutional Neural Networks, Contrastive Learning, Tumor Grading.

1. Introduction

Kidney cancer is a prevalent diagnosis, often identified at an advanced stage due to the lack of early symptoms (Pantuck et al., 2000). At this point, the disease may have metastasized, necessitating more aggressive treatments. For initial evaluation, diagnostic imaging methods such as computed tomography (CT) or magnetic resonance imaging (MRI) are commonly recommended to determine whether a tumor is malignant.

Benign tumors, though they may grow in size, do not spread to other organs and are generally not life-threatening. In contrast, malignant tumors pose a significant health risk. Malignant tumors are further classified into grades using systems like the Fuhrman (Lang

et al., 2005) or WHO/ISUP grading systems (Delahunt et al., 2019), with Grade 1 and Grade 2 categorized as low grade, and Grade 3 and Grade 4 as high grade.

However, accurately distinguishing between benign and malignant tumors, and assigning the correct grade, can be challenging (Frank et al., 2003; Johnson et al., 2015; Patel et al., 2016). While biopsies provide precise classifications, they are invasive and carry risks such as infection (Silverman et al., 2006). Consequently, there is a growing interest in developing non-invasive methods to assist doctors in pre-surgical evaluations.

Recent advances in contrastive learning have shown promise in various domains, including image classification and object detection (Hu et al., 2024; Wang et al., 2022; Xie et al., 2021). This approach is particularly robust in scenarios involving semi-supervised learning and data imbalance (Zhu et al., 2022; Chen et al., 2020). The core idea of contrastive learning is to maximize the similarity between an anchor and its positive sample, often created through data augmentation. However, applying this technique to CT images is not straightforward, which led us to introduce radiomic features as auxiliary knowledge augmentation.

Radiomics, which involves extracting a wide range of quantitative features from CT images, has become crucial in identifying tumor characteristics that are not visible to the naked eye (Gillies et al., 2016; Uhm et al., 2021). This process involves a sophisticated pipeline that processes medical imaging data (Mayerhoefer et al., 2020; Van Timmeren et al., 2020), with the features forming a rich dataset for training machine learning algorithms (Giraud et al., 2019; Rastegar et al., 2020).

In this paper, we introduce a novel model, the Radiomics-Guided Convolutional Neural Network (RGCNN), which integrates contrastive learning with both radiomic features and CT images. Our research shows that RGCNN can generate innovative features, termed CNN-based radiomics (CR). When combined with traditional radiomics (TR) as input for machine learning models, these features significantly enhance classification performance. We validated our model across three datasets, demonstrating its effectiveness in distinguishing between benign and malignant tumors, as well as in accurately classifying tumors by grade.

2. Related Works

2.1. Machine Learning Models Integrated with Radiomics

In recent years, machine learning and deep learning models have become increasingly prevalent in medical research. Specifically, several studies have integrated radiomic features into ML and DL models to predict classification outcomes in kidney tumors (Kocak et al., 2020; Sun et al., 2020). Uhlig et al. (2020) investigated the application of eight different ML algorithms to tackle the multiclass classification of kidney tumor subtypes, including chromophobe and papillary Renal Cell Carcinoma (RCC), Angiomyolipoma (AML), and oncocytoma. In their study, radiomic features with near-zero variance were excluded to eliminate linear dependencies within feature groups. The authors explored three feature selection strategies, combined with feature extraction through Principal Component Analysis (PCA) (Abdi and Williams, 2010). A 10-fold cross-validation approach was used, and class imbalance was addressed using the Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al., 2002). The model's performance was assessed using the Area Under the Curve (AUC) metric. Among the evaluated ML models, XGBoost achieved the highest

AUC values across all feature selection methods, with the best results obtained using no feature selection combined with SMOTE (AUC = 0.72).

2.2. CNN Integrated with Radiomics

The rapid advancement of technology has led to significant breakthroughs in deep learning models applied to CT images integrated with radiomics (Tez, 2025; Wang et al., 2025; Zhang et al., 2025). Lou et al. (2019) introduced a pioneering deep learning framework called Deep Profiler, designed to tailor radiotherapy doses by integrating CT images with radiomic features. The model consists of an encoder, a decoder, and a task-specific network, all working together to extract imaging features, estimate handcrafted radiomics, and predict therapy outcomes. The encoder utilizes a three-dimensional CNN to reconstruct new radiomic features and estimate the time to event for patients undergoing radiotherapy. By incorporating a proportional hazards model (Kumar and Klefsjö, 1994) and a nested five-fold cross-validation experiment, Deep Profiler demonstrated superior predictive capabilities compared to standalone deep learning and classical radiomics. A notable strength of this approach is its ability to identify salient voxels around the tumor, offering a significant improvement in personalized dose delivery by enhancing tumor volume delineation through automatic contouring. This integrated methodology, which combines imaging data with clinical variables, represents a significant advancement in improving treatment outcomes in radiotherapy. Han et al. (2022) introduced a novel framework for accurate classification and localization of abnormalities in chest X-rays, leveraging end-to-end semi-supervised knowledge-augmented contrastive learning. This approach excels in disease classification and localization while addressing the challenge of acquiring expensive annotated labels and bounding boxes. The framework uniquely integrates radiomic features as knowledge augmentation, creating a feedback loop where image and radiomic features reinforce each other. Experiments on the NIH Chest X-ray dataset demonstrate the framework's superior performance compared to existing baselines, advancing the application of contrastive learning in medical imaging.

3. Datasets

The datasets used in this study included KiTS (The 2019 Kidney and Kidney Tumor Segmentation Challenge (Heller et al., 2021)), TCGA-KIRC (The Cancer Genome Atlas Kidney Renal Clear Cell Carcinoma Collection (Linehan and Ricketts, 2019; Weinstein et al., 2013)), and a private dataset from Veterans General Hospital in Taichung (VGHTC), Taiwan. Both KiTS and TCGA-KIRC are public datasets, while the VGHTC dataset is proprietary. All CT images were taken preoperatively, prior to tumor removal surgery. The KiTS dataset contains 300 arterial phase abdominal CT scans of kidney cancer patients, along with ground truth semantic segmentations for the kidney, cyst, and tumor. For TCGA-KIRC, after excluding cases with missing labels and high-resolution CT images, we utilized 103 CT images. Since the TCGA-KIRC dataset lacked segmentation data, we applied a pre-trained nnUNet model (Johnson et al., 2015) from KiTS to generate the necessary masks. It is important to note that the TCGA-KIRC dataset exclusively contains CT images from malignant cases. The VGHTC dataset includes 215 CT images, selected from the first arterial phase scan of

each patient. Experienced, board-certified radiologists at VGHTC labeled the segmentation data and determined both the malignancy and grade of the tumors.

4. Methodology

4.1. Radiomics Guided CNN

Our goal was to seamlessly integrate information from 3D CT images and radiomic data, combining the spatial richness of the images with the detailed quantitative features derived from radiomic analysis. To achieve this, we utilized contrastive learning to create embeddings that capture the relationships between CNN-extracted features and radiomics data. This led to the development of a novel model, the RGCNN, which exploits the complementary strengths of both image-based CNN representations and radiomics-derived features to enhance predictive performance.

Our architecture includes two distinct neural networks: the image network and the radiomics network, as depicted in Figure 1. The image network includes an encoder, which is a 3D CNN with Deep Profiler inspired from Han et al. (2022) as its backbone (Figure 2), and a projector composed of a two-layer MLP with ReLU activation. The radiomics network, on the other hand, utilizes a three-layer MLP as its projector. Importantly, these networks and projectors operate independently, each with its own set of weights.

For the image view v, the encoder generates a representation $y_i = f_i(v)$, which is then projected to $z_i = g_i(y_i)$ through the image projector. Similarly, for the radiomics, a representation y_r is projected to $z_r = g_r(y_r)$ via the radiomics projector. We minimized the discrepancy between the corresponding projections z_i and z_r of the same patient using supervised contrastive loss (Khosla et al., 2020). Additionally, to handle classification errors from the representation y_i, we applied focal loss (Lin et al., 2017). During the testing phase, only the image encoder f_i is used.

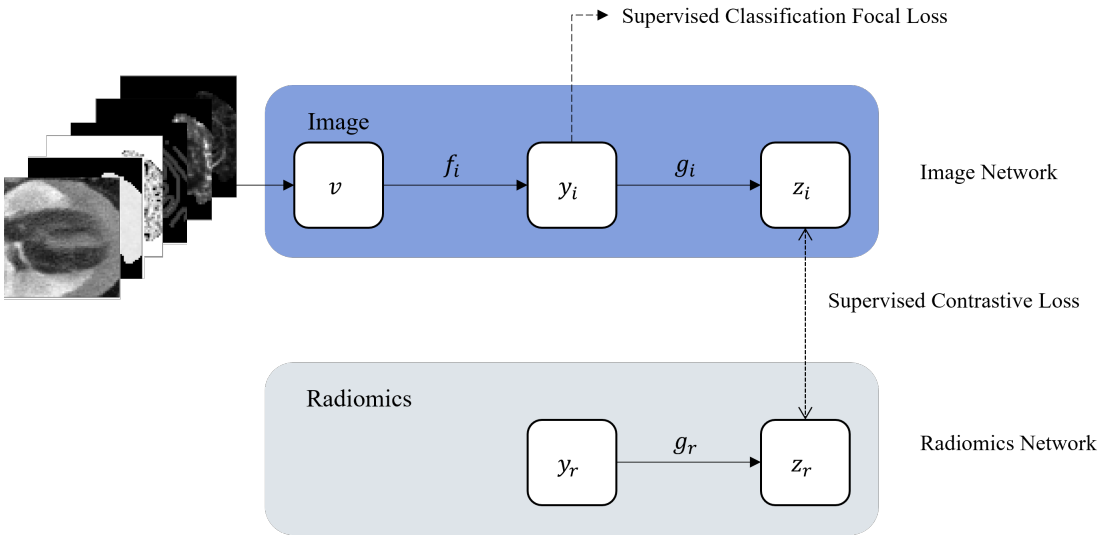


Figure 1: Overview of the RGCNN model

The image network in the RGCNN model takes 6-channel images as input, comprising the original cropped CT image with dimensions of 64×64×64 voxels, alongside five optimal radiomics 3D features. Meanwhile, the radiomics network is trained using the most effective radiomics 2D features to complement the image network. By adopting a contrastive learning approach, our model enhances its robustness, particularly in handling class-imbalanced data—a common challenge in medical imaging. In this framework, the positive samples are derived from pairing the image features and radiomic features of the same case, while negative samples are drawn from different-class images within the 6-channel dataset. To identify the most relevant 3D radiomics features, we conducted extensive CNN experiments, ultimately selecting the top five features for model input.

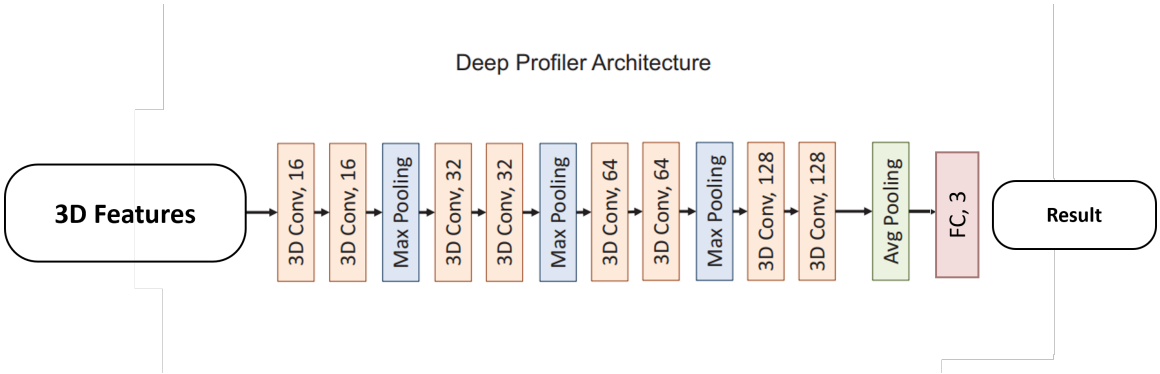


Figure 2: Model Structure of Deep Profiler

4.2. Loss Function

Overall learning loss function as Equation Equation (1) was formulated as a combination of supervised classification loss and supervised contrastive loss.

$$\mathcal{L}^* = \mathcal{L}^{fl} + \lambda \times \mathcal{L}^{cl} \tag{1}$$

Here, $\mathcal{L}^{fl}$ represents the supervised focal loss, $\mathcal{L}^{cl}$ stands for the supervised contrastive loss, and λ is the weighting factor controlling the importance of the supervised contrastive loss. The objective of supervised classification focal loss is to improve the model's performance by placing greater emphasis on errors and reducing the impact of predicting easy examples. The formulation of supervised classification focal loss $\mathcal{L}^{fl}$ is as follows:

$$\begin{aligned} \mathcal{L}^{fl} &= -(1-p_t)^{\gamma} \times \log p_t \\ p_t &= \begin{cases} p & \text{if } y = 1 \\ 1-p & \text{if } y = 0 \end{cases} \end{aligned} \tag{2}$$

p_t represents the predicted probability of the correct class, and $\gamma >= 0$ is a tunable parameter. y stands for the label, and p represents the predicted probability of the correct class at a given instance. The purpose of γ is to adjust the behavior of the focal loss. When

$\gamma > 0$, the loss is reduced for well-classified examples, enabling the model to focus more on difficult and misclassified instances. This flexibility allows the model to adapt its sensitivity to different types of errors according to the dataset's characteristics.

The supervised contrastive loss (L^{cl}) is designed to enhance the discriminative power of feature representations in a supervised setting. It encourages the model to bring representations of samples from the same class (positive pairs) closer in the feature space while pushing representations of samples from different classes (negative pairs) farther apart. The extended formulation for each pair-wise loss term (L_i^{cl}) is expressed as follows:

$$\begin{aligned} \mathcal{L}^{cl} &= \sum_{i=1}^{2N} \mathcal{L}_i^{cl} \\ \mathcal{L}_i^{cl} &= \frac{-1}{2N_{y_i} - 1} \sum_{j=1}^{2N} \mathbb{1}_{i \neq j} \cdot \mathbb{1}_{y_i \neq y_j} \cdot \log \frac{\exp(z_i \cdot z_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{i \neq k} \cdot \exp(z_i \cdot z_k)/\tau)} \end{aligned} \tag{3}$$

Here, L_i^{cl} represents the supervised contrastive loss for the i^{th} sample. The loss is computed over $2N$ samples in the batch. The indicator functions $\mathbb{1}_{i \neq j}$ and $\mathbb{1}_{y_i} \neq y_j$ ensure that the terms are only considered for different samples and same classes, respectively. The softmax-like operation in the numerator and denominator helps scale the similarity scores between representations z_i and z_j by the temperature parameter τ, controlling the sharpness of the distribution. The loss is then averaged over the number of valid pairs in the batch. Overall, the supervised contrastive loss encourages the model to learn robust and discriminative representations by focusing on the relationships between positive and negative pairs within the training batch.

4.3. CNN-Based Radiomics Features

In our study, we leveraged the feature maps generated by the RGCNN model to develop a novel set of features termed CNN-based radiomics (CR), inspired by traditional radiomics (TR) but designed to capture the rich spatial information and intricate patterns revealed by the RGCNN. Unlike TR, which applies predefined metrics directly to the Region of Interest (ROI) in tumor images, CR features are derived by summing the values within the feature maps. This method condenses the hierarchical representations learned across different layers into quantitative descriptors. By capturing the tumor's structural characteristics and the spatial relationships across multiple dimensions, CR features address aspects often overlooked by TR methods.

The integration of CR features aims to bridge the gap between the interpretability of TR features and the representational power of deep learning-based feature maps. By aggregating information from the feature maps, we harness spatial characteristics that contribute to a more comprehensive understanding of tumor complexities. The optimal outcome was achieved with 96 CR features, derived from layers 0 to 3, compared to 1316 TR features, highlighting the efficiency and potential of CR in enhancing tumor classification and assessment.

5. Experiments

5.1. Data Preprocessing

To standardize varying voxel spacing across diverse datasets, we performed isotropic resampling to 1.0×1.0×1.0 mm3 using B-spline interpolation. We then employed the Connected Component 3D (CC3D) (Silversmith, 2021) algorithm to isolate relevant tumor volumes. Utilizing 26-connectivity, we identified the largest tumor component (excluding volumes smaller than 100 voxels) and cropped the region accordingly; smaller tumors were fitted to a fixed (64,64,64) bounding box.

5.2. Radiomics Feature Extraction

We utilized the PyRadiomics package to extract features from CT images and masks. We applied original, wavelet (8 decompositions), and Laplacian of Gaussian (LoG, with $\sigma \in \{1.0, \ldots, 5.0\}$) filters. This pipeline generated 580 3D voxel-based feature maps and 1,316 2D scalar features, referred to as TR. These features encompassed first-order statistics, Gray Level Co-occurrence Matrix (GLCM), and Gray Level Run Length Matrix (GLRLM) metrics, capturing intensity distributions and textural patterns within the ROI.

5.3. Implementation Details

The dataset was partitioned into a 20% independent test set and an 80% training set, the latter subjected to 5-fold cross-validation with strict patient separation. We evaluated Random Forest (n_estimators=100, max_depth=10) (Ho, 1995) and XGBoost (Guestrin and Chen, 2016) classifiers. To mitigate overfitting, Principal Component Analysis (PCA) was employed to reduce both TR and CR features to 10 principal components. We addressed class imbalance using SMOTE for machine learning models and a "WeightedRandomSampler" for deep learning training. Model performance was quantified using the F1 Score and Area Under the Curve (AUC).

5.4. Result

We compared the performance of the proposed RGCNN model with several other approaches, including a standard CNN (architecture depicted in Figure 2), a TR-based classifier, a CR-based classifier, and the combined integration of CR and TR features.

Table 1 presents the classification results for distinguishing between benign and malignant tumors. Notably, the combination of CR and TR features resulted in significant improvements compared to using either CR or TR alone. The AUC score for the RGCNN model was the highest, followed closely by the CR+TR combination, demonstrating the effectiveness of our proposed model and the value of integrating CR features. While CR+TR outperformed in most metrics, it slightly underperformed in specificity, where the CNN-only model excelled. This suggests that the CNN model is particularly effective at correctly identifying benign cases but struggles more with malignant cases. On the other hand, TR ranked second in sensitivity, indicating its strength in identifying malignant tumors. Overall, the RGCNN model, by leveraging the strengths of both CNN and TR features, achieved the highest AUC, validating the robustness and effectiveness of our approach.

Table 1: Performance evaluation in benign and malignant task

	Benign vs Malignant				
	Accuracy	AUC	F1-Score	Sensitivity	Specificity
RGCNN	0.804 [0.780-0.830]	**0.831 [0.799-0.868]**	0.867 [0.847-0.853]	0.826 [0.8-0.853]	0.731 [0.678-0.798]
CNN	0.768 [0.741-0.796]	0.812 [0.776-0.852]	0.834 [0.809-0.855]	0.754 [0.722-0.789]	**0.816 [0.761-0.865]**
Radiomics (TR)	0.829 [0.806-0.854]	0.792 [0.752-0.843]	0.888 [0.871-0.906]	0.885 [0.865-0.909]	0.638 [0.584-0.717]
CR	0.777 [0.744-0.809	0.731 [0.698-0.782]	0.855 [0.833-0.882]	0.855 [0.826-0.883]	0.511 [0.455-0.600]
CR+TR	**0.830 [0.803-0.858]**	0.819 [0.788-0.856]	**0.891 [0.872-0.909]**	**0.901 [0.877-0.924]**	0.589 [0.528-0.664]

Table 2 presents the classification results for distinguishing between low-grade and high-grade tumors. Compared to Table 1, the overall scores are lower, indicating that this classification task is more challenging. Despite this, our results reveal a significant finding: the combination of CR and TR outperformed all other models across all metrics in this difficult task. This suggests that CR successfully captures critical features that might be overlooked by other methods. The strength of this combined approach lies in the complementary nature of the information each modality captures. TR, with its robust texture analysis, excels in identifying intricate patterns within the ROI. Meanwhile, CR, powered by deep learning, extends beyond the ROI, capturing valuable contextual information outside the explicit tumor boundaries. This synergy between TR and CR enhances the model's ability to distinguish between low-grade and high-grade tumors, highlighting the effectiveness of integrating traditional and deep learning-based radiomics features.

Table 2: Performance evaluation in low-grade and high-grade task

	Low-grade vs High-grade				
	Accuracy	AUC	F1-Score	Sensitivity	Specificity
RGCNN	0.717 [0.683-0.753]	0.716 [0.676-0.751]	0.613 [0.565-0.663]	0.589 [0.531-0.647]	0.796 [0.757-0.837]
CNN	0.695 [0.658-0.731]	0.710 [0.665-0.755]	0.592 [0.530-0.641]	0.583 [0.525-0.643]	0.763 [0.723-0.805]
Radiomics (TR)	0.708 [0.670-0.738]	0.734 [0.696-0.768]	0.610 [0.559-0.659]	0.601 [0.538-0.660]	0.774 [0.727-0.815]
CR	0.701 [0.663-0.731]	0.705 [0.668-0.743]	0.600 [0.551-0.648]	0.589 [0.532-0.647]	0.770 [0.730-0.808]
CR+TR	**0.740 [0.710-0.771]**	**0.737 [0.698-0.775]**	**0.637 [0.581-0.682]**	**0.601 [0.533-0.669]**	**0.825 [0.782-0.856]**

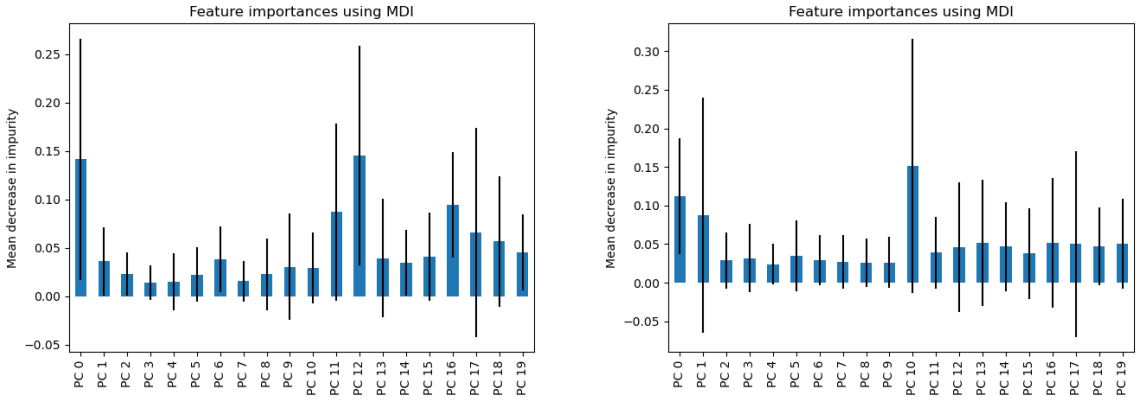
6. Discussion

6.1. Investigating the Importance of CR

By merging these diverse sources of information, we achieved a synergistic effect in Table 1 and Table 2. The texture-rich insights from TR seamlessly integrated with the spatial awareness and contextual understanding provided by CR. This could not only inherit the interpretability of TR but also leveraged the representational power of CR. Thus, we then wanted to further investigate how CR affect the overall performance.

The Mean Decrease in Impurity (MDI) is a measure used in Random Forest algorithms to assess the importance of each feature in making accurate predictions. Figure 3(a) and Figure 3(b) are the feature importances of the result from the Random Forest model on benign/malignant and high-grade/low-grade task respectively. Principal Component (PC) stands for the represented features after the operation of PCA. The first 10 features (PC0 $\sim$ PC9) came from CR and last 10 features (PC10 $\sim$ PC19) came from TR.

From Figure 3(a), the model's highest importance came from PC12 which was generated from TR, but there was also the contribution from PC0 which came from CR. This indicated that not only TR influenced the model, but CR features also assisted the model, enhancing its performance. As for the low-grade and high-grade tasks, in Figure 3(b), the model's highest importances were in PC10 and PC0. Similarly, the model was not biased towards one side, confirming that the CR indeed contributed to the performance of TR. This strategy emerged as a promising avenue for advancing tumor classification methodologies, showcasing the potential of amalgamating traditional and deep learning-based radiomic approaches for enhanced diagnostic accuracy.



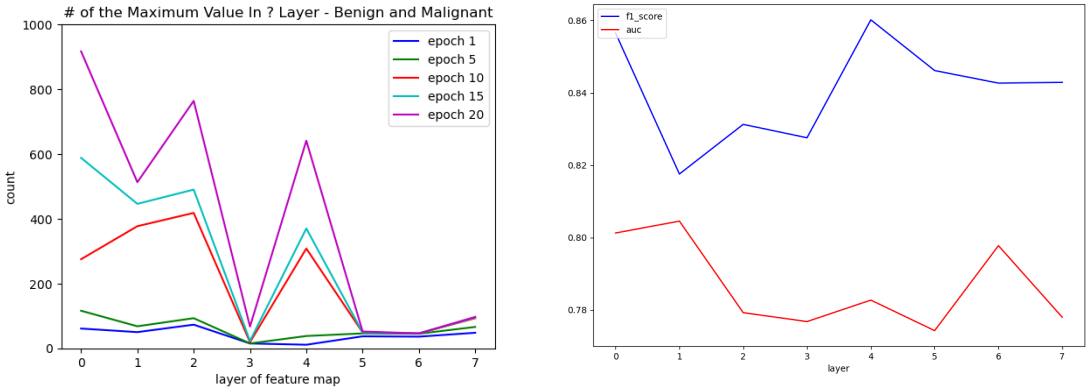
(a) Benign/Malignant Importances (b) Low/High Importances

Figure 3: Overview of Feature Importances

6.2. The Relationship Between Radiomics and Feature Maps

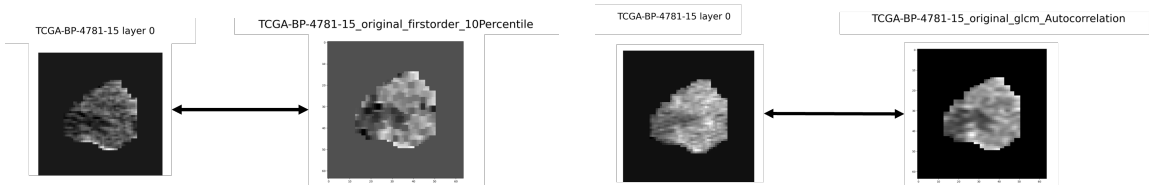
To quantify the correlation between radiomics features and feature maps across different layers, we calculated structural similarity index measure (SSIM) scores, as visualized in Figure 4. For the benign and malignant classification task, the optimal performance observed

at layer 0 coincided with the highest SSIM score. Additionally, a distinct peak was noted at layer 4, followed by a gradual decline. To further illustrate these relationships, the most correlated radiomics features are depicted in Figure 5.



(*a*) SSIM score between different layers (*b*) Performance of feature maps

Figure 4: The relationships between SSIM score and feature map performances



(*a*) original_firstoerder_10Percentile (*b*) original_glcm_autocorrelation

Figure 5: The most related radiomics features in layer 0.

7. Conclusion

This study integrates 3D CNN and radiomics to enhance the preoperative classification of kidney tumor malignancy and grade. By isolating the largest tumor segment, we analyzed feature maps to generate novel CR features. Our results demonstrate that combining CR with traditional TR significantly outperforms single-modality approaches. Future work will investigate layer-wise contrastive loss to strengthen the alignment between CNN representations and radiomic features, aiming for superior diagnostic accuracy.

References

Hervé Abdi and Lynne J Williams. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459, 2010.

Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16: 321–357, 2002.

Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey E Hinton. Big self-supervised models are strong semi-supervised learners. *Advances in neural information processing systems*, 33:22243–22255, 2020.

Brett Delahunt, John N Eble, Lars Egevad, and Hemamali Samaratunga. Grading of renal cell carcinoma. *Histopathology*, 74(1):4–17, 2019.

Igor Frank, Michael L Blute, John C Cheville, Christine M Lohse, Amy L Weaver, and Horst Zincke. Solid renal tumors: an analysis of pathological features related to tumor size. *The Journal of urology*, 170(6):2217–2220, 2003.

Robert J Gillies, Paul E Kinahan, and Hedvig Hricak. Radiomics: images are more than pictures, they are data. *Radiology*, 278(2):563–577, 2016.

Paul Giraud, Philippe Giraud, Anne Gasnier, Radouane El Ayachy, Sarah Kreps, Jean-Philippe Foy, Catherine Durdux, Florence Huguet, Anita Burgun, and Jean-Emmanuel Bibault. Radiomics and machine learning for radiotherapy in head and neck cancers. *Frontiers in oncology*, 9:174, 2019.

Carlos Guestrin and Tianqi Chen. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794. Association for Computing Machinery New York, 2016.

Yan Han, Chongyan Chen, Ahmed Tewfik, Benjamin Glicksberg, Ying Ding, Yifan Peng, and Zhangyang Wang. Knowledge-augmented contrastive learning for abnormality classification and localization in chest x-rays with radiomics using a feedback loop. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 2465–2474, 2022.

Nicholas Heller, Fabian Isensee, Klaus H Maier-Hein, Xiaoshuai Hou, Chunmei Xie, Fengyi Li, Yang Nan, Guangrui Mu, Zhiyong Lin, Miofei Han, et al. The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge. *Medical image analysis*, 67:101821, 2021.

Tin Kam Ho. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pages 278–282. IEEE, 1995.

Haigen Hu, Xiaoyuan Wang, Yan Zhang, Qi Chen, and Qiu Guan. A comprehensive survey on contrastive learning. *Neurocomputing*, 610:128645, 2024.

David C Johnson, Josip Vukina, Angela B Smith, Anne-Marie Meyer, Stephanie B Wheeler, Tzy-Mey Kuo, Hung-Jui Tan, Michael E Woods, Mathew C Raynor, Eric M Wallen, et al. Preoperatively misclassified, surgically removed benign renal masses: a systematic review of surgical series and united states population level burden estimate. *The Journal of urology*, 193(1):30–35, 2015.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673, 2020.

Burak Kocak, Ece Ates Kus, Aytul Hande Yardimci, Ceyda Turan Bektas, and Ozgur Kilickesmez. Machine learning in radiomic renal mass characterization: fundamentals, applications, challenges, and future directions. *American Journal of Roentgenology*, 215 (4):920–928, 2020.

Dhananjay Kumar and Bengt Klefsjö. Proportional hazards model: a review. *Reliability Engineering & System Safety*, 44(2):177–188, 1994.

Hervé Lang, Véronique Lindner, Marc de Fromont, Vincent Molinié, Hervé Letourneux, Nicolas Meyer, Mael Martin, and Didier Jacqmin. Multicenter determination of optimal interobserver agreement using the fuhrman grading system for renal cell carcinoma: assessment of 241 patients with¿ 15-year follow-up. *Cancer*, 103(3):625–629, 2005.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

W Marston Linehan and Christopher J Ricketts. The cancer genome atlas of renal cell carcinoma: findings and clinical implications. *Nature Reviews Urology*, 16(9):539–552, 2019.

Bin Lou, Semihcan Doken, Tingliang Zhuang, Danielle Wingerter, Mishka Gidwani, Nilesh Mistry, Lance Ladic, Ali Kamen, and Mohamed E Abazeed. An image-based deep learning framework for individualising radiotherapy dose: a retrospective analysis of outcome prediction. *The Lancet Digital Health*, 1(3):e136–e147, 2019.

Marius E Mayerhoefer, Andrzej Materka, Georg Langs, Ida Häggström, Piotr Szczypiński, Peter Gibbs, and Gary Cook. Introduction to radiomics. *Journal of Nuclear Medicine*, 61(4):488–495, 2020.

Allan J Pantuck, Amnon Zisman, Mitchell K Rauch, and Arie Belldegrun. Incidental renal tumors. *Urology*, 56(2):190–196, 2000.

Hiten D Patel, Michael H Johnson, Phillip M Pierorazio, Stephen M Sozio, Ritu Sharma, Emmanuel Iyoha, Eric B Bass, and Mohamad E Allaf. Diagnostic accuracy and risks of biopsy in the diagnosis of a renal mass suspicious for localized renal cell carcinoma: systematic review of the literature. *The Journal of urology*, 195(5):1340–1347, 2016.

Sajjad Rastegar, M Vaziri, Younes Qasempour, MR Akhash, Neda Abdalvand, Isaac Shiri, Hamid Abdollahi, and Habib Zaidi. Radiomics for classification of bone mineral loss: a machine learning study. *Diagnostic and interventional imaging*, 101(9):599–610, 2020.

Stuart G Silverman, Yu Unn Gan, Koenraad J Mortele, Kemal Tuncali, and Edmund S Cibas. Renal masses in the adult patient: the role of percutaneous biopsy. *Radiology*, 240(1):6–22, 2006.

William Silversmith. cc3d: Connected components on multilabel 3d & 2d images. *Zenodo*, 2021.

Xue-Ying Sun, Qiu-Xia Feng, Xun Xu, Jing Zhang, Fei-Peng Zhu, Yan-Hao Yang, and Yu-Dong Zhang. Radiologic-radiomic machine learning models for differentiation of benign and malignant solid renal masses: comparison with expert-level radiologists. *American Journal of Roentgenology*, 214(1):W44–W54, 2020.

Mesut Tez. Deep learning radiomics: Redefining precision oncology through noninvasive insights into the tumor immune microenvironment. *World Journal of Gastrointestinal Oncology*, 17(7):108175, 2025.

Johannes Uhlig, Andreas Leha, Laura M Delonge, Anna-Maria Haack, Brian Shuch, Hyun S Kim, Felix Bremmer, Lutz Trojan, Joachim Lotz, and Annemarie Uhlig. Radiomic features and machine learning for the discrimination of renal tumor histological subtypes: a pragmatic study using clinical-routine computed tomography. *Cancers*, 12(10):3010, 2020.

Kwang-Hyun Uhm, Seung-Won Jung, Moon Hyung Choi, Hong-Kyu Shin, Jae-Ik Yoo, Se Won Oh, Jee Young Kim, Hyun Gi Kim, Young Joon Lee, Seo Yeon Youn, et al. Deep learning for end-to-end kidney cancer diagnosis on multi-phase abdominal computed tomography. *NPJ precision oncology*, 5(1):54, 2021.

Janita E Van Timmeren, Davide Cester, Stephanie Tanadini-Lang, Hatem Alkadhi, and Bettina Baessler. Radiomics in medical imaging—"how-to" guide and critical reflection. *Insights into imaging*, 11(1):91, 2020.

Dong Wang, Huan Liu, Zhuo Cao, Weiqian Huang, Wen Fu, Li Shao, Ji Zhang, Wanyu Su, Xianwen Yu, Ce Han, et al. Integrating deep learning and radiomics in the differentiation of major histological subtypes of invasive non-mucinous lung adenocarcinoma using positron emission tomography and computed tomography. *Translational Lung Cancer Research*, 14(9):3323–3336, 2025.

Xiyue Wang, Sen Yang, Jun Zhang, Minghui Wang, Jing Zhang, Wei Yang, Junzhou Huang, and Xiao Han. Transformer-based unsupervised contrastive learning for histopathological image classification. *Medical image analysis*, 81:102559, 2022.

John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M Stuart. The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113–1120, 2013.

Enze Xie, Jian Ding, Wenhai Wang, Xiaohang Zhan, Hang Xu, Peize Sun, Zhenguo Li, and Ping Luo. Detco: Unsupervised contrastive learning for object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8392–8401, 2021.

Zhen Zhang, Tianchen Luo, Meng Yan, Haixia Shen, Kaiyi Tao, Jian Zeng, Jingping Yuan, Min Fang, Jian Zheng, Inigo Bermejo, et al. Voxel-level radiomics and deep learning for predicting pathologic complete response in esophageal squamous cell carcinoma after neoadjuvant immunotherapy and chemotherapy. *Journal for immunotherapy of cancer*, 13(3):e011149, 2025.

Jianggang Zhu, Zheng Wang, Jingjing Chen, Yi-Ping Phoebe Chen, and Yu-Gang Jiang. Balanced contrastive learning for long-tailed visual recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6908–6917, 2022.